

WHAT IS CLAIMED IS:

1. A system to assist a user in classifying documents to concepts, the system including a user interface device, including an output device configured to provide a user at least one term from a document and corresponding relevance information indicating whether the term is likely related to at least one concept, the user interface device also including an input device configured to receive from the user first assignment information indicating whether the term should be assigned to the at least one concept for classifying documents to the at least one concept.
2. The system of claim 1, further including a document classifier, the document classifier including an input receiving the documents and the concepts, and including an output providing at least one classification of at least one of the documents to at least one of the concepts, the document classifier including instructions to be executed to classify the at least one document to the at least one concept by comparing terms in the documents to user-assigned terms assigned to the concepts.
3. The system of claim 2, further including a knowledge map including multiple taxonomies, each taxonomy including at least one concept node representing a particular concept.
4. The system of claim 1, further including a candidate features extractor, the candidate feature extractor including an input receiving the documents, the candidate feature extractor including an output, which is coupled to the user interface, the extractor output providing candidate terms from the documents from which the user can select at least one term to be assigned to at least one concept.
5. The system of claim 1, in which the user interface input device is also configured to receive from the user second assignment information indicating whether at least one document should be assigned to at least one concept for

extracting terms from the at least one document from which the user can select at least one term to be assigned to the at least one concept for classifying documents to the at least one concept.

5    6.     The system of claim 1, in which the output device includes a taxonomy display listing taxonomies for which at least one term and corresponding relevance information is available.

7.     The system of claim 1, in which the output device includes a concept node  
10 display listing concept nodes for which at least one term and corresponding relevance information is available.

8.     The system of claim 1, in which the output device includes a term display listing at least one term and corresponding relevance information.

15    9.     A method of assisting a user in classifying documents to concepts, the method including:

providing a user at least one term from a document and corresponding relevance information indicating whether the term is likely related to at least one  
20 concept; and

receiving from the user first assignment information indicating whether the term should be assigned to the at least one concept for classifying documents to the at least one concept.

25    10.    The method of claim 9, further including assigning or deassigning the term to the at least one concept using the first assignment information, to provide at least one user-assigned term corresponding to the at least one concept.

11.    The method of claim 10, further including classifying documents to concepts  
30 by comparing terms in the documents to the at least one user-assigned term.

12. The method of claim 11, further including computing the relevance information using results from the classifying documents to concepts.

13. The method of claim 9, further including forming multiple taxonomies for  
5 organizing concepts.

14. The method of claim 9, further including receiving from the user second assignment information indicating whether at least one document should be assigned to at least one concept.

10

15. The method of claim 14, further including extracting terms from the at least one document from which the user can select at least one term to be assigned to the at least one concept for classifying documents to the at least one concept.

15 16. The method of claim 9, further including providing the user information about taxonomies for which at least one term and corresponding relevance information is available.

17. The method of claim 9, further including providing the user information  
20 about concept nodes for which at least one term and corresponding relevance information is available.

18. A system to assist a user in classifying a document, in a set of documents, to at least one node, in set of nodes, in a taxonomy in a set of multiple taxonomies, the  
25 system including:

a candidate feature extractor, including an input receiving the set of documents and an output providing candidate features extracted automatically from the document without human intervention;

a user-selected feature/node list, including those candidate features that have  
30 been selected by the user and assigned to nodes in the multiple taxonomies for use in classifying the documents to the nodes;

a user interface, to output the nodes and candidate features, and to receive user-input selecting and assigning features to corresponding nodes for inclusion in the user-selected feature/node list; and

a document classifier, coupled to receive the user-selected feature/node list,  
5 to classify the documents to the nodes in the multiple taxonomies.

19. The system of claim 18, in which the document classifier includes:  
a first input receiving the set of documents;  
a second input receiving the user-selected feature/node list;  
10 a third input receiving multiple taxonomies; and  
an output providing, edge weights from the documents to the nodes.

20. The system of claim 18, in which the user interface outputs, for a document  
selected by the user, those features corresponding to that particular document.  
15

21. The system of claim 18, in which the user interface outputs, for a document,  
a corresponding indicator of how successfully the document classifier classified the  
document to the nodes in the multiple taxonomies.

22. The system of claim 21, in which the user interface outputs a list of the  
documents ranked according to the number of nodes to which each document was  
classified by the document classifier.  
20

23. The system of claim 18, in which the user-interface outputs a representation  
25 of the multiple taxonomies.

24. The system of claim 18, in which the document classifier includes a first  
input receiving a selected subset of the set of documents, each document in the  
subset assigned by the user to at least one node, and in which the document  
30 classifier classifies the set of documents to nodes in the multiple taxonomies using  
features of the selected subset of documents.

25. A method including:  
extracting automatically candidate features from a set of documents;  
outputting to a user an indication of the candidate features;  
5 outputting to the user an indication of relevance of the candidate features to  
nodes;  
receiving user input of user-selected features and user-assignments of the  
user-selected features to nodes; and  
classifying documents to nodes in multiple taxonomies using the user-  
10 selected features and corresponding user-assignments.
26. The method of claim 25, further including providing, for each document,  
those features corresponding to that particular document.
- 15 27. The method of claim 15, including outputting an indication of how  
successfully a document was classified.
28. The method of claim 27, in which the outputting includes providing a list of  
the documents ranked according to the number of nodes to which each document  
20 was classified.
29. The method of claim 25, further including outputting a representation of the  
multiple taxonomies.
- 25 30. The method of claim 25, further including receiving user input of a user-  
selected subset of the set of documents, and wherein the receiving user input of  
user-selected features and user-assignments of the user-selected features to nodes is  
performed on features obtained from the user-selected subset of the documents.